Locating place names from place descriptions

Article *in* International Journal of Geographical Information Science · December 2013

DOI: 10.1080/13658816.2013.785550

3 authors:

Maria Vasardani
University of Melbourne

Stephan Winter
University of Melbourne

Kai-Florian Richter
Umeå University

Locating Place Names from Place Descriptions

Maria Vasardani, Stephan Winter, Kai-Florian Richter

Department of Infrastructure Engineering, University of Melbourne

Parkville, 3010, VIC, Australia

{mvasardani, winter, kfrichter}@unimelb.edu.au

In this paper we review current literature on geographic information retrieval based on place names. We focus on the positional uncertainties and extent vagueness frequently associated with place names in linguistic place descriptions and on the differences between common users' perception and the way that geographic information services interpret place names. We argue that despite some notable efforts from the scientific community, geographic information services still cannot unambiguously recognize and sufficiently reason spatially with place names from linguistic expressions. We focus on three interrelated research areas: (a) the use of place names in gazetteers, (b) the use of formal models to reason with spatial relations and with the spatial extent of place names in linguistic place descriptions, and (c) Web harvesting and crowd sourcing techniques for identifying place names and their spatial extension from public and volunteer sources such as social networks and photo-sharing sites. We identify some opportunities for synthesizing existing approaches that would expedite the process of intelligent communication about place names between services and users. We discuss the shortcomings of the current state of affairs in locating place names from place descriptions, and identify new areas of importance for future research.

Keywords: geographic information retrieval, place names, georeferencing, gazetteers, spatial reasoning

1. Introduction

In everyday communication, people often use place names and spatial relations to describe where they are, to give navigational instructions, to inform of the locations of events, and to generally convey spatial information that is based on the shared knowledge of these names. Place names also frequently occur in text documents with geographic context. The rising popularity of geospatial services, including navigation services, trip guides, mobile location based games, and check-in social networks, expose machines more and more to human communication patterns. Also, the Web as a resource of text-based information that represents human geographic knowledge necessitates more intuitive methods for the automated identification of localities described by place names (Jones *et al.* 2001, Schlieder *et al.* 2001, Hill 2006, Twaroch, Smart, *et al.* 2008). Place names provide a qualitative reference system with which geographic locations are described (Liu *et al.* 2009), though sometimes these are ambiguous (Longley *et al.* 2010). While place names are first-class citizens in both the conceptual and associative structure of cognition as well as in language, most geographic information systems and services are developed on unambiguous, crisp, and metric geometries removed from human concepts. These systems are yet to exhibit user interfaces that properly interpret linguistic place descriptions (Wang

1994, Kuhn 2001). It is this fundamental mismatch, causing conflicts in the interface between human and machine, this paper focuses on.

Services that provide geographic information using place names, such as navigation and routing systems, transportation timetables, environmental forecasting, map-based websites, and Web search engines, typically utilize gazetteer-based resources to identify the localities and resolve ambiguities in the use of names for which multiple instances exist. Gazetteers are depositories of georeferenced geographic names, structured by a taxonomy, and storing the text (name), the type, map coordinates usually in the form of a representative point, and some organizational hierarchy, such as county, region, state, or nation associated with each name (Hill *et al.* 1999, Hill 2006). Authoritative gazetteers are often derived from national map series, reflecting the official or administrative view of geographic space (Jones *et al.* 2008), and similar gazetteers exist for points of interests, or for addresses. Crowd-sourced gazetteers also exist, for example *geonames.org* or *openaddresses.org*. If everybody were satisfied with the current available official representation of places, then place modelling would not be an issue (Davies *et al.* 2009, Winter and Truelove 2010). However, there are frequent instances where national and regional governments, emergency services, disaster prediction and management services, health services, trip planners or navigation systems fail, confuse or frustrate people. This is because the name or assumed extent of a particular place (a neighbourhood, address, landmark or facility) that people often have in mind, is not the same as the official (Meegan and Mitchell 2001, Davies *et al.* 2009). Winter and Truelove (2010) have collected some examples of problems that arise from simple local orientation and local searches in Web-mapping services.

The present paper offers a survey of the existing literature contributing to locating place names. We focus on the positional uncertainties and extent vagueness frequently associated with place names and with the differences between common users' perception and the representation of places in gazetteers. We believe that even though a single unified solution might not yet exist, from examining the current compartmentalized solutions, we can offer an insight into the interplay between them and the potential benefits of a synthesized approach.

The remainder of the paper is structured as follows: In Section 2 we will define place descriptions, based on place names, to define the scope of the paper. Then, we will examine place names from three interrelated areas: In Section 3 we will examine the cataloguing of place names in gazetteers. We will review different ways of georeferencing place names in gazetteers, and how the use of geographical ontologies in gazetteers has been suggested to compensate for the lack of semantics and context dependent specifications attached to place names, and for the interoperability among different gazetteers. In Section 4 we will study models from formal logic based qualitative spatial reasoning that have been employed to extract the location corresponding to place names in place descriptions. We will also examine how probabilistic and membership based techniques have dealt with the issues of uncertainty and vagueness of the place extents that names refer to. In Section 5, we will examine different attempts to identify the location of lesser-known or vernacular names from their relation to better known or defined places from crowd sourcing (human descriptions) and from Web harvesting, and how such techniques are used to enrich traditional gazetteers with more place names.

Through this review, in Section 6 we offer an educated opinion about the current state of qualitative geographic information retrieval based on place names. We also raise some issues that have not yet been fully addressed, and outline future research opportunities for locating place names.

2. Identifying place descriptions

In this paper, we analyse the existing literature on extracting the geographic location of a place *l* from linguistic place descriptions which resemble answers to the question "Where is *l*?" For 'place' we adopt the general definition provided in Bennett and Agarwal (2007): "Places are the conceptual entities that enable cognitive structuring of the spatial aspects of reality". We consider place descriptions as locative expressions used to describe a scene where the locatum *l* is located with respect to reference object(s) (Dini and Di Tomaso 1995). Such place descriptions comprise a referring expression (RE), an optional verb, and a set of prepositional phrases (PPs):

$$[\text{RE}]\ [(\text{opt})\ \text{verb}]\ [(\text{PP}_1, \text{PP}_2 \ldots \text{PP}_n), n \geq 1] \qquad (1)$$

For example, the place description 'The village is to the south of the ski hill, near the highway exit' consists of the RE 'the village', a verb 'is', and two PPs: the first one is 'to the south of the ski hill', and the second one is 'near the highway exit'. In general, the RE—the place name referring to the locatum *l*—may be a proper name ('London'), a NP of any kind (for example revealing *type*, as in an indefinite NP—'a hotel'—or a definite NP—'the station'), or an anaphoric NP ('it'). The reference object(s) captured in the PPs are place names themselves. The preposition in each PP introduces a spatial relation between the spatial extensions of the locatum and the place referred to by the place name in the PP. Thus, any place description can be formally represented by:

$$\text{PD} = \{\ r_i\,(\text{ext}(l), \text{ext}(g_i)), i \in (1..n)\ \} \qquad (2)$$

where r_i is the spatial relation of the i^{th} PP, ext(l) is the spatial extension of the locatum l, and ext(g_i) is the spatial extension of the reference object g_i in PP_i (Bennett and Agarwal 2007). The verb can elaborate further details of the spatial relations (Dini and Di Tomaso 1995). This representation reveals that the cognitive notion of 'place' comprises both its linguistic and its spatial extensions (Fig. 1), and that there must be a mapping ext() from the linguistic extension to the spatial extension.

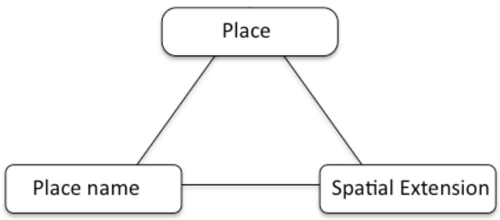

Figure 1: The (cognitive) concept of place, its linguistic extension, and its spatial extension.

From the definition of the place descriptions considered here, a few difficulties related to locating a place *l* from natural language descriptions are immediately obvious. For example, previous research has provided formal models for the qualitative spatial relations denoted by the prepositions in the PPs. Such models define topological, distance, and absolute or relative direction relations. However, these formal models are too strict to fully capture human cognition and natural languages' flexibility. An example for cognitive flexibility is the context-dependency of qualitative distance relationships such as 'near'; an example for language flexibility is the ambiguity of some of these relations such as 'at' that can have interpretations as in being 'in', 'in touch', or 'near' a place. Such ambiguities may be resolved only when considering the context of the conversation. While context related interpretations are clearly important and will be part of the discussion, this study excludes any discourse analysis and concentrates instead on the use of place names and relations found in verbal place descriptions in various methods for locating *l*.

Another difficulty related to locating *l* is the frequent disagreement in the spatial extension of places. By 'locate' we mean: *to determine or specify the spatial extension of the locatum, in relation to other already located references, or spatial reference systems*. Accordingly, the operation ext() can have various flavours, such as returning a relative or absolute spatial extension. Next, we will see how researchers have so far dealt with place names and their spatial extensions in gazetteers.

3. Places in digital gazetteers

In lieu of semantically annotated place descriptions as they are promised by the semantic Web, most Web search engines use gazetteers for the identification of place names in spatial queries and for the generation of the place names' spatial extensions. Gazetteers are an important part of enterprise georeferencing systems, Web-mapping services, navigation services, geographic information retrieval (GIR), and the semantic Web. They also provide an indexing framework for geoparsing—the recognition of references to geographic locations in the form of place names in text strings, and the subsequent assignment of geographic identifiers to them.

Gazetteers consist of structured information about place names, linking the names of places to thematic concepts, and to the representation of their spatial extension according to a mathematical framework, i.e., a geographic footprint (Hill 2000, Goodchild and Hill 2008) (Fig. 2). In this sense, gazetteers can be regarded as "geospatial dictionaries of geographic names" (Hill 2000) and are, therefore, commonly used for information retrieval based on place names. Specifically, by linking place names to coordinates and thematic data, gazetteers enable the retrieval of indirectly georeferenced information (e.g., place names, postal codes), the vertical data integration in the form of attribute data about the same place, facilitate integration of online information systems, and efficiently handle very large quantities of geographic place names (Schlieder et al. 2001).

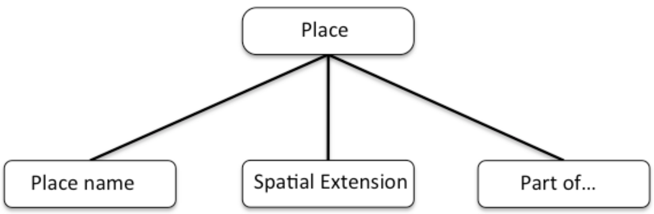

Figure 2: The gazetteer data model of place names.

When it comes to gazetteer performance, the nature of each of the gazetteers' core components—names, categories or types, and geographic locations—may create a number of problems. For example, the place names included in gazetteers are the subset of names in use that are proper or authoritative. Gazetteers do not usually include unofficial, vernacular names. The characteristics of included place names are also undefined, they are often not unique (i.e., there exist multiple locations with the same name), nor do they always uniquely identify a location (i.e., there are locations with multiple names, maybe varying by date, spelling, language, or official vs. colloquial use) (Goodchild and Hill 2008). Names may also be assigned only for specific purposes, and used only in specialized domains. In addition, the feature type associated with each place name is selected from a gazetteer-specific taxonomy or ontology, which is not shared between gazetteers, and most often not adopted in common language place descriptions. Some of these issues were identified in a specialist meeting, which brought together researchers and practitioners who were working with the definition, modelling and application of gazetteers and gazetteer services (Goodchild and Hill 2008). Even though short- and long-term research and development directions were identified in that meeting, not all of these have been pursued equally in the post-meeting period. It seems more weight has been put on the enrichment of existing gazetteers (developed from official sources) with place names used in colloquial language, and on the creation of a geographic ontology umbrella, under which the various feature typing schemas can be combined for allowing gazetteer interoperability. These are important issues, and are discussed in more detail in Sections 3.2, 4 and 5, where formal models and Web mining techniques are employed for the enrichment of gazetteers.

However, more issues have been identified which are of equally significant importance for the efficient functioning of gazetteers, but have been somewhat neglected so far. For example, the combination of different spatial representations in the same gazetteer, which would enable the generation of query results on different levels of granularity, has not yet been explored to a substantial extent.

3.1. Footprints of places

Geographic footprints may be single points of geographic coordinates, bounding boxes or similar approximate geometries, or other polygons describing the shape of the entity to finer degrees of detail. For example, gazetteers typically store geographic coordinates with a place name, originally for text labelling purposes. In many contexts point locations suffice for representing entities, for example for disambiguation (Hill 2000). However, for performing spatial reasoning operations such as testing for overlap or *part-of* relations places need to be represented with certain extents. We will later discuss forms of these representations, such as crisp or non-crisp boundaries (Section 4), which are required to cope with our limited ability to draw boundaries around many places (Smith 1995, Smith and Varzi 1997). The ability to commit to boundaries, and the degree of detail these boundaries can be given, directly impact any spatial reasoning. While for some operations points or minimum bounding rectangles are enough, at least in first iteration, more elaborate and thus more costly representations may be needed for other operations (Hill 2000).

Regarding gazetteer footprint representations, Schlieder *et al*. (2001) aim to combine the advantages of detailed polygon representations with the computational efficiency and ease of less detailed representations. They do so by abstracting the partonomies that result from recursively applying the *part-of* relation to planar polygonal tessellations, into connection graphs. The nodes in such graphs correspond to the polygons of the tessellations, and edges between them denote the binary *part-of* relations. The connection graph is basically the neighbourhood graph of the tessellation together with its combinatorial embedding in the plane and can, thus, support multiple neighbourhood relations between two polygons, as well as a number of graph-theoretical operations. Such operations can use the topological and ordinal information encoded in the connection graph for making inferences to answer expressive spatial queries. For example, if the initial query results are not satisfactory, such operations can expand the query to the area containing the query footprint, or to its neighbouring footprints, as these are computed from the graph. The authors' experiments demonstrate better query results when using the extended query footprint in combination with the ontology of their gazetteer.

The connection graph approach is a promising starting point for constructing a linking framework of geographic representations of different dimensions and detail in gazetteer-based services. For example, at the bottom level of the graph, the first iteration of the *part-of* relation could be applied for linking polygon centroids with individual polygons, so that if the query was such that point representations would be enough, the service could return them, instead of the individual polygons.

3.2. Ontologies of places

Researchers have argued for the benefits of enriching gazetteers with terminological reasoning based on ontologies of feature types (e.g., Fu *et al.* 2005) and with spatial reasoning capabilities for the retrieval of better query results. Apart from techniques to measure geographical neighbourhood and closeness (see Section 4.1), efforts focused on methods to evaluate the semantic closeness of place names to the topic of interest in GIR results, especially where there is an imprecise match between the query place name and the candidate place names. To this end, Jones *et al.* (2001) suggest an ontology of place that may be used to derive semantic distance measures for enriching gazetteers in GIR. The ontology comprises topological and proximity relations between places, and sparse coordinate data found in gazetteers that represent the geographical footprint of places. The places are classified into geographical categories and are linked to non-geographical concepts classified in conceptual hierarchies. It is assumed that a place is distinguished by the sum of the geographical regions, or other parent places to which it belongs, either directly or by hierarchical inheritance. For example a town may be part of the formal hierarchical administrative subdivision of a nation, and also belong to a physiographical hierarchy based on prominent landscape features, such as mountain ridges or peaks. The hierarchical distance measure is combined with the Euclidean distance between centroids of places, resulting in a hybrid spatial distance measure. This measure is itself combined with a thematic distance (based on classification semantics) to form an integrated semantic closeness measure that can rank GIR results for relevance.

The development of this ontology introduces the interesting idea of combining both spatial distance and thematic distance as measures of similarity between candidate place names in query results. In the context of GIR, this idea connects Tobler's (1970) first law of geography ("…near things are more related than distant things") with thematic resemblance and similarity. As we will see next, semantic similarity measures become necessary tools for the successful integration and interoperability of different gazetteers.

3.2.1. *Interoperability of place gazetteers*

A vision shared by many in gazetteer research is the development of a distributed local-responsibility service network of collaborating gazetteers, instead of an all-in-one global gazetteer (e.g., Janowicz and Keßler 2008). The main idea is that in such a network, each gazetteer offers lookup for places within its spatial and thematic scope, and higher level gazetteers distribute the queries appropriately. For this idea to materialize, apart from resolving issues of mismatching names and geographic footprints or different representations altogether among gazetteers (Janee 2006, Hastings 2008), the variety of taxonomies that gazetteers use to categorize places creates the biggest interoperability hurdle for cross-gazetteer access.

Janowicz and Keßler (2008) propose a feature type ontology based on the Alexandria Digital Library’s Feature Type Thesaurus (FTT) which gazetteers can use to improve their interoperability and their reasoning capability. Conceptually, the level of abstraction in the ontology is at the same time generic enough to be used and further refined by multiple gazetteers, and detailed enough to enable direct type-lookup. It is, according to Guarino's (1998) distinction, a *domain* ontology into which each gazetteer can add supplementary sub-concepts to describe their own perspective, while due to the shared and formal vocabulary of the ontology itself, type lookups can span over many gazetteers. The authors also integrated *semantic similarity* and *subsumption* based reasoning into the shared feature type ontology. Semantic similarity is about the comparison of meanings of the types concepts. The idea behind subsumption based retrieval, as described by Lutz and Klien (2006), is to return results about all sub-concepts of the specified search concept—based on a *is-a* or a subclass-of relationship—that satisfy the user's requirements (Janowicz *et al.* 2011). Specifically, they use the SIM-DL measure to evaluate the semantic similarity of feature types in the proposed ontology (Janowicz 2006, Janowicz *et al.* 2007). Similarity between concepts is measured by comparing their definitions for overlap, with a value of 1 indicating that two compared concepts are identified by the same superconcepts, while 0 means the concepts have nothing in common. SIM-DL combines both subsumption and similarity reasoning and supports high-level description logics. Janowicz and Keßler (2008) also suggest an extended gazetteer Web interface which covers both search functionality and presentation of results, allowing users to continuously refine their queries until desirable results are achieved.

The methodologies of place name representations and gazetteer interoperability discussed so far are based on typical gazetteers, developed from some official authority. The rest of this paper, however, takes also into consideration place names as they are used in everyday linguistic place descriptions, reflecting people’s cognitive models of places. For example, we next discuss a proposed method with which gazetteers can be enriched with place names form the Web.

3.2.2. *Public knowledge in place gazetteers*

Keßler, Janowicz, *et al.* (2009), when referring to next generation digital gazetteers, suggest the incorporation of information about places from implicit and explicit volunteered geographic information (VGI) with appropriate harvest and extraction methodologies. Their methodology for identifying possible place name candidates is based on creating spatio-temporal clusters of place names mentioned in tags of volunteered information. While, in Section 5 particularly, we investigate various techniques developed to extract the spatial extensions of place names identified in Web text documents and tagged online photo media, these approaches rather work with a predefined set of place names without considering its expansion. The approach in Keßler, Janowicz, *et al.* (2009), however, recognizes the need for an automated way of identifying new place names and enriching gazetteers with (especially vernacular) place names found in VGI. For example, the idea that the photo tags of possible place name candidates will cluster spatially, but will have uniform temporal distribution works well in two ways: (1) it identifies place name candidates, and (2) it introduces the concept of 'events' such as wildfires and demonstrations, in the sense of temporary places, also discussed in more detail in (Keßler, Maué, *et al.* 2009). This new idea poses the interesting question as to whether places where events of dynamic nature happen should be considered equally to the more permanent place names.

Next, we cover examples from literature that focus on the use of formal logic to deal with the notion of vagueness in place descriptions. In particular, both the footprints of place names, and the spatial relations between them may be vague (e.g., 'north of the Rocky mountains') in which case special considerations are necessary for the retrieval of valid and sufficient geographic information based on place names (Goodchild and Hill 2008). Formal logic based methods laid the foundations for most subsequent crowd sourcing and Web harvesting techniques.

4. Qualitative Spatial Reasoning for Place Descriptions

Spatial relations are commonly used in place descriptions. People generally perceive the geographic world through embodied experiences, graphical, or narrative descriptions. Their mental spatial representations are configurational, hierarchic, and known to be distorted (Hirtle and Jonides 1985, Tversky 1993). Accordingly people prefer to formulate or answer spatial queries using qualitative descriptions, rather than numeric measures such as the ones stored or computed in geographic databases. Qualitative spatial relations also play an important role in qualitative spatial reasoning (QSR), the logical reasoning about spatial environments in artificial intelligence (Cohn and Hazarika 2001), reflecting cognitive capabilities of people for computational efficiency in applications such as robotics. QSR is also an essential concept in Egenhofer and Mark's (1995) Naïve Geography, an effort to cater for the non-expert geographical views and to support people's cognitive geographic models in geographic information systems and science. Spatial relations within Naïve Geography are treated from two complementary sources: (1) the cognitive and linguistic approach, and (2) the formal approach, which can be implemented in computers and is based on mathematical models. There is ample literature for both approaches (for the first, see for example Talmy 1983, Herskovits 1985, Retz-Schmidt 1988, Landau and Jackendoff 1993, Grabowski and Weiss 1996, Vorwerg and Rickheit 1998, Levinson 2003), but in this section we concentrate on the formal approaches for reasoning with spatial relations and the spatial extents of places.

4.1. Formal models of spatial relations

There are well-defined formalizations in QSR of topological, directional, and distance relations, or combinations thereof. For topological relations there are two main approaches of formal models, one based on point-set topology (Egenhofer and Herring 1990, Egenhofer and Franzosa 1991, Clementini *et al.* 1993, Shariff *et al.*

1998, Chen *et al.* 2003), and the second based on regions in space as the primitive objects and connectedness as the primitive relation (Randell *et al.* 1992, Bennett 1994, Cohn 1995, Cohn *et al.* 1997). Formal models of qualitative distance and proximity relations can be found for example in Sharma *et al.* (1994), Gahegan (1995), and Hernández *et al.* (1995), for direction relations in Ligozat (1993), for cardinal directions in Freksa (1992), for combined distance and direction relations in Frank (1992) and Clementini *et al.* (1997), and for order and projective relations in Hernández (1991), Schlieder (1995), Clementini and Di Felice (1997, 2000), Eschenbach and Kulik (1997), and in Clementini (2011).

From early on, researchers in QSR have compared formal models of qualitative spatial relations with cognitive models with which people communicate about places using natural language spatial terms. From human subjects experiments for example, Mark and Egenhofer (1994) conclude that the 9-interection method has potential for defining cognitively meaningful natural language spatial predicates, but needs to be supplemented with finer metric details. Worboys *et al.* (2004) conducted experiments with human subjects on conceptions of proximity, and direction and angle. The results suggest a context-dependent relationship between commonsense notions of proximity and geometric distance, while commonsense notions of directions exhibit strong similarities with the geometric models of direction. Results also point to the reasonable assumption that people's proximity and direction judgements are derived from a common mental representation; therefore, combinations of proximity and direction formal models could be more efficient for qualitative spatial reasoning. From experiments with human subjects on the role that major axes (e.g., the perpendicular left and right) play on the conceptualization of turn directions, Klippel and Montello (2007) concluded that a combination of sectors and axes, such as the double cross calculus in Freksa (1992), is a sensible approach.

Even though such experiments show a relatively good correspondence between certain formal models of relations and human cognitive models, the fact that most formal models are limiting the expression of relations to a predefined set (for example 8 topological relations in the 9-intersection model or 4 relations with a single cross for directions) makes the success of such comparisons limited. Linguistic place descriptions are much richer in expressions; thus, better comparison surveys and techniques are needed for verifying whether and how expressional needs of natural language can be covered with formal models of finite sets of relations.

Success in that area would also guarantee better results for efforts such as Yao and Thill's (2006) conceptual framework. Their work focuses on closing the gap between the way that metric spatial information systems answer spatial queries and the general public's naïve geographic view of the environment. Their framework equips a geographic information system with a language interpreter and a qualitative spatial reasoner. While the traditional information system provides the tools and data for conventional spatial data handling, the language interpreter parses a user query into place names, qualitative spatial relations, measures, numbers, logical operators, and translated fuzzy linguistic expressions and sends the qualitative spatial relations (if there are any) to the qualitative spatial reasoner. If no such relations are found, the query is handled in the information system. The reasoner comprises many qualitative reasoning modules based on the aforementioned formal theories, appropriate for handling each type of relation (e.g., topological, directional, proximity, etc.). The goal of such a conceptual framework is to enhance traditional spatial information systems that rely mostly on quantitative information, with the capability to handle qualitative spatial queries expressed in natural language.

Studies demonstrate that cognitive concepts of spatial relations may differ from formal concepts (e.g., Klippel and Montello (2007)). Also the cognitive concepts rely on the domain of discourse (Worboys 2001), which is not captured in formal concepts. Finally, in contrast to formal models people seem to interpret a spatial expression over its contrast, e.g., north in contrast to south (Tenbrink and Freksa 2009) and human reasoning happens over such opposites, at least in a preferential manner (Schleipen *et al.* 2007). Thus, more evidence is necessary for proving that existing formal models of spatial relations adequately depict people's relational representations, or richer models are needed for better depicting such representations.

4.2. Models of vagueness and uncertainty for spatial relations

Qualitative spatial relations, such as 'near', 'not far from', or 'to the southeast of' used in place descriptions can be vague, conforming to the existence of borderline cases and susceptible to the sorites paradox (Williamson 1994). Quantitative spatial relations on the other hand can suffer measurement and precision uncertainties, both because of the devices and of the human operating them. Next, we review literature focusing on models that deal with vagueness and uncertainty of spatial relations.

4.2.1. Vagueness

Robinson (1990) used a question-answer approach with human subjects to determine the meaning of 'near', and used 'distance' as the surrogate variable. The system 'learns' what 'near' means by constructing fuzzy sets of places near the reference place(s). Dutta (1990, 1991) developed an integrated framework for approximate reasoning with imprecise, incomplete, and possibly conflicting quantitative and qualitative relations. He used fuzzy logic as the computational basis for both representing quantitative information and interpreting linguistically expressed qualitative constraints. Worboys (2001) investigated how formal theories of vague spatial relations, and specifically three-valued logic, or equivalently 'egg-yolk' (Lehmann and Cohn 1994, Cohn and Gotts 1996) or broad boundary regions (Clementini and Di Felice 1996) can be applied to the concept of nearness. He tested data from human subjects for significance and provided a set of 'nearness neighbourhoods' of selected places in a university campus environment, which remain dependent on the domain of discourse. Regions with broad boundaries provided a good first approximation for these regions, but not without limitations (e.g., necessity for crisp boundaries between 'near, 'indeterminate', and 'not near' areas). Du et al. (2004) provided membership functions for cardinal direction relations, arguing that fuzzy logic can better describe cardinal direction relations.

4.2.2. Uncertainty

In order to deal with the uncertainties associated with descriptions of point localities (e.g., locality data in museum collections), Wieczorek et al. (2004) developed a method which combines different kinds of locality uncertainties and measurement imprecision into a single maximum error. The method is called 'point radius method'. It describes localities using a circular area centred at the coordinate pair assigned to the locality, with a radius that is calculated after combining six different sources of uncertainty or imprecision: extent of the locality, unknown datum, map scale, and imprecise distance, direction and coordinate measurements. This method was refined in Guo et al. (2008) by: (a) considering the actual shape of the reference object from which the measurements are taken, and (b) positioning the locality on a probability surface of the probability distribution of all the different uncertainty sources. Liu et al. (2009), working on a related method, developed the concept of an uncertainty field—a two-dimensional probability density function for representing an uncertain locality from topological, directional and metric spatial assertions.

The work by Liu *et al.* (2009) is a significant effort to combine the various qualitative and quantitative vagueness and uncertainties when dealing with place descriptions. The downside is that the proposed approach focuses on descriptions about point objects. However, according to Bennett and Agarwal (2007), descriptions, such as 'inside the South of England', can be viewed as expressions for a place as well, corresponding to a (vague) region. Therefore, a reasonable expectation that has not yet been addressed is the expansion of methods for determining the spatial extension of places, which will consider qualitative vagueness and quantitative uncertainties of spatial relations. In the next subsection, we examine some methods developed to deal with vague spatial extensions.

4.3. Formal models of vague place extents

In natural language place descriptions, 'places' correspond to mental representations of the locations of conceptualized spatial features. These locations lack precise boundaries since people do not always agree with the boundaries imposed on places for satisfying the needs of authoritative organizations, or with each other's idea about the exact location or extent of a place, as is often the case with vernacular place names. Past literature includes formal models of regions with vague or indeterminate boundaries. Some of the techniques employed for creating such regions are the 'egg-yolk' with concentric sub-regions that indicate degrees of membership in a vague region (Lehmann and Cohn 1994, Cohn and Gotts 1996), broad boundary regions (Clementini and Di Felice 1996), rough sets (Bittner and Stell 2000, 2003), fuzzy sets (Fisher 1996, Fonte and Lodwick 2004, Dilo *et al.* 2007), supervaluation (Kulik 2001) and granular partitions (Bittner and Smith 2001). Acknowledging the need for a model that preserves the imprecision or vagueness of a location and does not force any precise, fuzzy, or 'rough' approximations, Galton and Hood (2005) introduced anchoring, a method which uses the topological relations of imprecise or vague locations to regions whose boundary is more precise, providing vague locations with a reference within a spatial data model. In Hood and Galton (2006) the authors implemented the method as a locational component for a spatial data model.

As a side product of a study about the user requirements of geographic information systems (GISs), Davies *et al.* (2009) discovered that most users would like any place data model to be able to satisfy two basic needs: (1) the need to find locations from place names, essentially answering the question 'Where is X located?', and (2) the need to find appropriate place names for locations, including local vernacular names. Most of them also expressed the desire to know the extent of the location that a place name refers to and agreed that this extent will be vague in most cases. Not all of the aforementioned techniques for modelling vague regions would be of equal value to all users, for all cases, however, and Montello *et al.* (2003) argued that the method depends also on the type of collected data. In a more extended analysis of the appropriateness of vague regions modelling techniques, Davies *et al.* (2009) highlight some remaining questions concerning the actual implementations of data models that deal with place names. As we will see in the next section, many Web harvesting methods for extracting the spatial extensions of place names from documents and other media sources use one of the aforementioned methods for depicting the vague regions on probabilistic and density surfaces.

5. Defining places with Web harvesting techniques

Place names are the references people use to specify a location by name. As place names are used within human communication, people use them also when accessing several types of public information systems and services, such as routing systems, transportation timetables, map-based websites and general Web search engines. Different studies have verified that considerable percentages (e.g., 14% - 18%) of user queries submitted to different online search engines contain at least one geographic-related term such as a place name, or adjectives indicating localities, or spatial relations (Sanderson and Kohler 2004, Delboni *et al.* 2007). Geographic queries can also be written using Equation 2 (Sec. 2), with l now the unknown place searched for.

To answer such a query usually means generating the geographic footprint (a geographic representation) of the place name the query refers to, modifying it according to the spatial relation, and using it as the reference framework for the query's subject. The perspective taken by administrative authorities often does not coincide with the extents people have in mind when using place names in natural language. Vanity addressing is an example of conflicting perspectives. There are also plenty of vague or vernacular place names, such as the 'Highlands of Scotland', the 'The Midwest', or the 'The Alps', for which even authorities have no definite extent.

Researchers recognize the need to acquire people's perception of vague and vernacular place names for enriching gazetteers and geographically focused services with approximate boundaries for such place names. Such an effort has already started in the SPIRIT project, where using a structured query interface, users can get answers, for example, about "hotels in the Scottish Highlands" (Jones *et al.* 2002).

Empirical data acquired by interviews with people are valuable for exploring cognitive extents of vague places. Montello *et al.* (2003), for example, asked people to draw on a map their perception of 'downtown Santa Barbara', first assuming a crisp boundary exists, and then adding 100% and 50% certainty boundaries. Clough and Pasley (2010) showed photos to people and asked them to identify whether depicted places, for which they knew the geographic coordinates, belong inside or outside of vague places such as the 'Sheffield City Centre'. An approximation of the area corresponding to the vague region was then calculated using Kernel Density Estimation (KDE) based on the points identified as within the region from the responses.

5.1. Extracting place names from the Web

The collection of empirical data, however, is tedious, time-consuming, and prone to scalability issues. Therefore, efforts have concentrated on using the Web as a readily available and rich source of semi-structured text documents that contain geographical information such as place names, addresses, postal codes, or phone numbers. Geographic information retrieval (GIR) is more complex than traditional information retrieval (IR), which focuses on accessing Web documents that are relevant to user queries submitted as a phrase or set of keywords. Central to GIR is linking language to space with the process of toponyms (or place names) resolution (Leidner 2007). In it simplest form, GIR performed in existing geographical search facilities typically uses gazetteers to match recognized place names with a geographic location.

Toponym resolution comprises geo-parsing and referent disambiguation. Geo-parsing is the process of detecting place names and associated spatial language qualifiers and distinguishing the geographical occurrences of place names usage from those that refer to other entities (e.g., is it Washington the place or the person?), thus, resolving *semantic* ambiguity (Jones and Purves 2008, Leveling and Hartrumpf 2008). Geo-parsing is often considered a special case of Named Entity Recognition (NER), which is a standard part of Natural Language Processing (NLP). However, a detailed analysis of NLP is out of the scope of this paper.

Once a name has been established as a geographic one, there is still the need of identifying a single unambiguous place it refers to (e.g., Athens, Greece or Athens, Georgia), also known as resolving *referent* ambiguity (Overell and Rüger 2008). For example, Stokes *et al.* (2008) studied NLP-based approaches for detecting and disambiguating place names and their annotation within documents. Using a manually annotated dataset to evaluate some standard NLP methods, they found that—for their experiment—simply employing a gazetteer to recognize the presence of place names outperformed the NLP methods.

There are various approaches for solving referent ambiguity. Simple rule-based disambiguation methods use special gazetteers with a single location per place name, which has been selected on various criteria such as size, population and relative importance (Li *et al.* 2003, Clough *et al.* 2004). More complex methods define a geographic scope for a document, by either checking where the document was published or applying a minimum bounding box around the locations appearing in the document, assuming that the referred locations are geographically close (Rauch *et al.* 2003, Amitay *et al.* 2004), while others use heuristic rules (Leidner *et al.* 2003). Contextual information in the two-to-five words preceding and following a place name can provide the geographic location type referred to or imply relationships with other locations (e.g., 'Athens, Georgia') (Rauch *et al.* 2003, Amitay *et al.* 2004, Clough *et al.* 2004). Data driven methods apply machine-learning to match place names to locations (e.g. Grossman and Frieder 2004) and hybrid approaches such as Bucher *et al.* (2005) apply semi-supervised techniques. Other studies look into co-occurrence models—sets of document models, where each model is an ordered list of place names as they occurred in the document, used for disambiguation (Overell and Rüger 2008). Co-occurrence models are richer than conventional gazetteers because they record the order in which place names occur within multiple documents. Hence, rules of association between place names can be constructed, based for example on resources such as Wikipedia.

5.2. Defining vague place extents from Web documents.

In the previous section we briefly discussed methods for acquiring place names from the Web. Frequently, users employ informal, vernacular place names, often without exact boundaries (e.g., 'the South of France'). The focus of this section is on harvesting information about vaguely defined spatial extensions of place names from the Web.

5.2.1. Acquiring an initial set of well-defined places

A common Web-harvesting technique for collecting documents that mention place names of vaguely defined regions is to look for other place names that frequently occur in association with the more extensive vague region, under the assumption that such places can be expected to be spatially related with the vague region. Such methods typically employ trigger phrases and other Web queries to reveal places that are contained, or lie in the vicinity of the vague region (e.g., Purves et al. 2005, Arampatzis et al. 2006, Jones et al. 2008). Identified places are collected as points (pairs of geospatial coordinates) that are in some relation to the vague region. From these points, an approximate boundary can be placed and calculated by various methods. For example, from knowledge of points that are defined as inside or outside, a boundary may be generated from interpolation. Both Voronoi diagrams and Delaunay triangulations have been adopted as solutions for generating the boundary (Gold et al. 1996, Alani et al. 2001, Arampatzis et al. 2006). In Schockaert et al. (2005), apart from knowledge of points that lie inside the vague region, additional information comes in the form of constraints, such as "*l* is located in the south of *g*". Here, *l* is a known place that is used to delineate the possible extent of *g*, a place of vague extent, by refining initial representations of it. Since these constraints are vague themselves, possibility distributions are proposed. Moreover, fuzzy belief revision techniques are used in order to discard the inevitably inconsistent constraints that arise from extensive Web searches.

If the points in relation to the vague region also have probabilities of membership attached to them, probability density surfaces may be generated through interpolation (Purves *et al.* 2005). Slices of such regions correspond then to sharp boundaries that include points with probabilities above a certain threshold value. Each of the recognized place names can also be weighted according to its occurrence frequency in the Web-search results. By exploiting statistical evidence for the inclusion of points within a vague region, the variation in confidence for the included points may be reflected (Jones *et al.* 2008). In addition, the generated density surface is itself a representation of the vague or vernacular place name. Twaroch, Jones, *et al.* (2008) used a similar KDE for representing the extend of vernacular place names mined from Gumtree and Google's Business Directory and Community Maps.

5.2.2. Other Web harvesting methods

The aforementioned techniques for deriving representations of imprecise and vague places rely on Web searches that result from a sufficiently large and representative initial set of points that lie within the region. And while it is safe to assume that such sets will exist for popular tourist destinations and city environments, especially when considering the recent increase in geotags that appear in social media, such an assumption falls short for places such as residential and local neighbourhoods, among others. There is then a need for additional information.

To that end, Vögele et al. (2003) extended the work in Schlieder et al. (2001) (see Section 3.1) and provided a method to add place names with vague spatial extensions to the connection graph of already defined related places. Information for the new place names comes from harvested information about the hierarchical relations between the added place and existing places. Based on assumptions made from research on preferred spatial mental models, the method provides lower and upper approximations for the extent of the added place names. The lower approximation comprises regions definitely inside or equivalent to the imprecise place, while the upper approximation comprises the lower approximation plus all overlapping regions. This method provides a possible means for reasoning with multiple place names, the partonomic relations of which to well-defined places are only partially defined. From a data storage point of view, we believe that the connection graph is an efficient way of storing the *part-of* hierarchical relations between polygonal representations and will also be fast enough for recall operations in spatial queries. The downside however, is that it is a limiting method, in that it can only store information about *part-of* relations (in essence only containment topological relations), excluding other types of spatial relations that may be harvested from the Web. We argue that different methods that take into consideration the plethora of information about spatial relations between places available on the web, such as the ones examined next, have better chances of defining spatial extensions comparable to those in people's cognitive models.

Schockaert *et al.* (2008) suggested a heuristic approach for deriving supplementary topological relations—containment and adjacency relations—between geographic regions, from implicit information in Web documents. Automatically extracting spatial relations from Web documents is a challenging problem, not least due to the sparseness of explicitly mentioned relations. Therefore, Schockaert *et al.* suggest to initially value high recall over high precision, and then to filter the extracted information about topological relations through a fuzzy spatial reasoner.

A recent study by people of the same group (Schockaert *et al.* 2011) suggests a technique for combining qualitative and quantitative information from heterogeneous spatial information for generating appropriate polygons for places. The problem of defining a region then becomes a combinatorial search problem that treats all available spatial information as constraints on allowed instantiations (polygons) of variables (regions), and which discretizes the plane appropriately. One of the main findings in this study is that taking into account available qualitative information in the form of relations, significantly improves the results of the KDE that is applied on the initial set of points that are found from Web queries to be in inside the vague region. Finally, introducing possibility distributions, Schockaert (2011) suggests that a way to model the uncertainty underlying vague regions is to attach a probability distribution to the set of possible delineations, as these derive from aforementioned techniques. By doing so, one can infer how likely each delineation is, thus creating a possibility distribution of the vague region delineations.

We believe that a synthesis of existing approaches would vastly improve the results of the Web harvesting approach. For example, the aforementioned method by Schockaert *et al*. (2011) could benefit from using a first polygonal approximation of places as this is defined in a connection graph like the one proposed in Vögele *et al*. (2003). This polygonal approximation could later be enriched with additional qualitative and quantitative spatial relations of different kinds between place names, as these are harvested from the Web. The advantage of such a synthesis is that the first approximation to a spatial extension of a place name can be found in a readily available connection graph of mereotopological hierarchies and become more refined with more detailed information harvested from Web documents. In addition, since the connection graph is based on preferred mental models' rules, it improves the possibility of creating spatial extensions about place names that are better match with people's perception of them. Another possible synthesized approach would be to combine the possibility distributions of vague regions' delineations that model uncertainty in Schockaert (2011) with the uncertainty fields in Liu et al. (2009) (Section 4.2.2), which consider different kinds of uncertainties. Synthesized approaches from already available methodology could speed up the creation of reliable methods for enrinching gazetteers and geographic services with a number of colloquial and non-autnoritative place names. In the next subsection, we will also examine different sources of vernacular place names.

5.3. Looking for places in other media

Apart from Web searches to discover place names in Web text documents, research has also focused on other data sources such as georeferenced photographs on photo sharing sites such as Flickr[1] and Panoramio[2] for discovering locations associated with a given place name. Flickr for example has been extensively used in research because of the provision of direct links between the place or object depicted and the geographic coordinates automatically acquired from GPS devices on phones and cameras, or later matched manually by the users, the dates of the photo taking, and the tags—textual descriptions of the photo subjects (Martins 2011).

Rattenbury and Naaman (2009) focus on spatial patterns in tags from Flickr. Specifically, they attempt to automatically determine if there is a coherent place with which each tag is associated. In this sense, the determination of the tag's place becomes the same as defining the dominant place from Web search queries (Wang *et al.* 2005). Rattenbury and Naaman's hypothesis is that the spatial distribution of the tags' usage is strongly geographically localized, therefore, a specific spatial region may be identified within the area of the tags spatial distribution. In other words, the spatial patterns of place tags should appear as bursts—a highly peaked probability distribution over a small number of nearby values (Jaffe *et al.* 2006). Rattenburry and Naaman adopt specific baseline methods, based on bursts detection techniques. These techniques measure the statistical significance of the occurrence of tags inside over outside of a specific region. Issues of autocorrelation and scale dependence of the tags usage spatial distribution are accounted for. By employing statistical significance tests at large enough scales, where the localization of tags is significantly stronger than on other scales, the authors suggest that tags can be associated with the regions formed from these localizations.

Even though they do not consider the temporal distribution of tags, as Keßler *et al.* (2009) do, we believe that this method has a lot of potential for valid identifications of vernacular place names. This is especially the case because of the consideration of scale and spatial autocorrelation, which can otherwise influence the results with false positives. Again, synthesizing the two clustering approaches for a method that combines the advantages of both would most likely prove to be beneficial.

In a different study (Grothe and Schaab 2009), the automatic delineation of vague regions based on Flickr photographs is developed with the use of two methods, one based on KDE, and another based on One-Class Support Vector Machines (Munoz and Moguerza 2006). Martins (2011) presents a refined method that takes as input a probability density surface created from Flickr's photo tags with methods such

[1] http://www.flickr.com

[2] http://www.panoramio.com

as Grothe and Schaab's (2009), and combines it with land coverage images, the pixels of which are assigned a value that corresponds to terrain features (e.g., vegetation, urban areas, deserts, water). The improvement of the vague regions' estimates is due to the fact that terrain discontinuities are introduced in the delineation of the regions' extent (e.g., a vague region is more likely to correspond to land, thus water pixels should not be included). Martin's method also removes some of the outliers in the set of points acquired from Flickr. We argue that this novel method, which stems from combining different approaches of spatial distribution of relevant information (e.g., photo tag probability density surfaces with remote sensing images) is promising, especially for the automatic identification of the types of place names, which can then be input in digital gazetteers.

All the above studies, however, do not account for inaccuracy and imprecision in the georeferencing and tagging of Flickr photographs, nor for the prominence of vernacular place names, and the use of generic versus specific vernacular place names, in particular. In a fashion similar to Liu *et al.* (2009) who studied various kinds of qualitative and quantitative uncertainty and imprecision of place descriptions in Web documents, Hollenstein and Purves (2010) focused on similar issues on descriptions collected from tagged photographs. Their study results show that overall, the precision and accuracy of user georeferenced and tagged images are sufficient for describing city neighbourhoods, for example, even though errors in both semantics and geotagging occur. In addition, more than 70% of georeferenced photographs and 35% of tags included at least one place name at some granularity level, with most of the vernacular place names being specific, rather than generic. The authors also used density surfaces for generating what proved to be reasonable representations of the regions associated with the photo-tags and demonstrated that the general public's overall behaviour toward creating metadata is sufficient for the generation of sub-city level footprints, for practical purposes (e.g., to use then in information systems as convex hulls or bounding boxes for querying purposes).

6. Discussion

In this paper we examined three interrelated areas of current research on the use of place names in linguistic place descriptions and in spatial queries posed to online services. In the first part, we reviewed some of the recent advances in cataloguing place names in digital gazetteers. Then we examined some of the formal logic based approaches for dealing with spatial relations between place names and for locating vague spatial extensions. Finally, we described examples of Web harvesting and crowd sourcing techniques for recognizing vernacular place names and defining their spatial extensions.

As we have already argued with examples throughout the review, synthesis of some of the approaches would be beneficial for empowering gazetteers and geographic services beyond the current state of the art. Also, each research area has associated issues, which we argue have not been researched to a satisfying extent. For example, many of the requirements for the new age of digital gazetteers were already identified in the specialist meeting on Digital Gazetteer Research & Practice in 2006 (Goodchild and Hill 2006), however, they failed to attract sufficient follow-up attention. We examined the issues that have mostly been researched since then, namely the enrichment of gazetteers with colloquial place names, and common ontology-based feature taxonomies for seamless interoperability among gazetteers.

Still, other topics, such as enabling gazetteers to connect place names with spatial extensions of different dimensionality, are equally important for providing users with spatial query results at different granularities, in appropriate contexts.

Context itself is another problematic factor that has not been fully resolved towards the automatic interpretation of natural language place descriptions. Regarding the interpretation of the use of spatial relations, for example, Gahegan (1995) did experimental work on qualitative measures of *proximity* and pointed out that the existence of connection paths between places, the scale, and objects' attractiveness are some of the important contextual factors that people use for judging nearness. For example, 'near' in the city scale can be expected to be of shorter distance than in the country scale, and cognitive anchor points distort or disturb symmetry (Couclelis *et al.* 1987). Geographic services are not yet really capable of identifying context factors, such as appropriate scales of a spatial query or what the user means by 'near'.

Cai *et al.* (2003) and Cai (2007) identified different contextual factors; those of task, spatial contexts and background of the user. We argue that the current methods of parsing linguistic descriptions with place names are not sufficient for accommodating for different user backgrounds and personal preferences. Even if services disregard 'personal' place names (i.e., those only known to a very small group of people), which arguably may not have a place in gazetteers, many current services allow queries on vague regions, such as 'the south of France' or 'the inner suburbs'. Most such services, however, cannot unambiguously handle issues of individual differences in understanding (the extent of) these regions. For example, while a specific suburb may be part of the inner suburbs for some, it may not be for others. Such individual differences may introduce a need for ranking, or alternatively for personalization in how a query is handled by a service, for example, by dedicated adaptations of vague place concepts.

Furthermore, even if services are equipped with formal calculi for dealing with the qualitative nature of relations (e.g., the ability to handle vague relations such as 'near', 'south of', 'far from'), these formal models are often not comparable with the cognitive models people use for spatial reasoning. For example, while the complement of 'north' in the formal set of four directional relations {north of, south of, east of, west of} is the set of the remaining three relations, Ragni *et al.* (2007) showed that 'not north' is mostly interpreted by people as its directional opposite, 'south of'. In other words, negation is not always interpreted by formal logic and people's logic in the same way.

Related to the problem of spatial relations' interpretation by geographic services is that of relationship synonyms. For example, while people have no problem handling synonyms such as 'near', 'close to' and 'not far from' according to the appropriateness of the situation, not all formal frameworks are expressive enough to distinguish between such subtleties.

Linguistic synonyms pose yet more problems for current gazetteers. Many places change names (and often extent too) with the passage of time. How many and which of the place names should be linked with the same place in gazetteers? If the place name associated with a certain spatial extent is not used anymore, does that place seize to exist? That also brings again into focus the question about dynamic events, such as those examined in Keßler, Janowicz, *et al.* (2009): they happen at certain locations and adopt a name for the period of time they happen. But do place names of dynamic events such as demonstrations, country-side festivals, or floods deserve a place in the gazetteers alongside place names of more permanent features?

Researchers have also pointed out that not only the flexibility of natural language, but different languages themselves pose problems for the ubiquitous use of digital gazetteers. David Mark, as a participant to the 2006 specialist meeting about research related to gazetteers (Goodchild and Hill 2006), gave a few examples of disparities between the types that can be associated with the same place names in different languages. For example, for the Navajo, some place names are associated with types for which in English no single word exist, such as “A canyon wall receiving sunlight”. They also have many geographic features for which at least two different proper names exist: the traditional, and the every-day one. Even between languages of countries of closer cultural proximity, such as English and French, the search by type for the same geographic features is not easy (e.g., ‘lake’ labeled features in English would not necessarily be categorized as ‘lake’ in French, but often as what is translated in English as ‘lagoon’ or ‘pool’). Such issues create serious challenges for the creation of geographic services at a global level, and the research on such issues is still in its initial stages.

Challenges for the creation of global geographic services are also presented by the fact that the Web harvesting and crowd sourcing techniques that were examined in Section 5 are methods that rely on the collection of data from a crowd, and with sufficient sampling for their information retrieval methods. Contributions from a crowd seems to rely on a certain demographic group which has the knowledge of and access to the technology necessary for creating the sources of information about place descriptions and sharing them on the Web. We would argue, therefore, that there are certain biases that are associated with the information collected from social media by this specific group of people that have the means to travel and leave evidence of their spatial tracks. There is a plethora of place descriptions in other languages, which span parts of the world not covered by the demographics mentioned above. This information, then, is not accessible for crowd sourcing techniques. Otherwise, it seems a promising approach to capture human understanding of place and place names.

This review aims to identify how close we are to a truly intelligent communication between geographic services and their human users, using place names to ask or answer geographic queries. From the discussion, we can safely argue that no such service yet exists, which could unambiguously and with no semantic or context related problems communicate about place names with users, similar to the ability of a human communication partner. As stated earlier, a unified solution is not yet entirely possible, which among others, would mean that formal models of geographic services resemble closely the cognitive models of users. However, we reviewed notable efforts on three different interrelated aspects of the problem, and believe that synthesis of some approaches is feasible, if not necessary for solving some of the visited issues. Such synthesis would provide major improvements.

Further, we identified additional challenges about the use of place names in natural language place descriptions, the resolution of which would take us closer to a possibility for truly universal, human-like geographic services that would intelligently communicate with people about their every-day spatial needs.

Acknowledgements: We wish to thank Matt Duckham, Lesley Stirling and Lawrence Cavedon for the valuable feedback and comments on various drafts of this paper. This work was funded by the ARC under its Linkage Scheme (LP100200199 “Talking about Place”).

References

Alani, H., Jones, C.B., and Tudhope, D., 2001. Voronoi-Based Region Approximation for Geographical Information Retrieval With Gazetteers. *International Journal of Geographical Information Science*, 15 (4), 287–306.

Amitay, E., Har'El, N., Sivan, R., and Soffer, A., 2004. Web-a-where: geotagging web content. *In*: *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 273–280.

Arampatzis, A., Kreveld, M.J. van, Reinbacher, I., Jones, C., Vaid, S., Clough, P., Joho, H., and Sanderson, M., 2006. Web-based Delineation of Imprecise Regions. *Computers, Environment and Urban Systems*, 30 (4), 436–459.

Bennett, B., 1994. Spatial Reasoning with Propositional Logics. *In*: *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*. Morgan Kaufmann, 51–62.

Bennett, B. and Agarwal, P., 2007. Semantic categories underlying the meaning of "place". *In*: *Proceedings of the 8th International Conference on Spatial Information Theory*. Berlin, Heidelberg: Springer-Verlag, 78–95.

Bittner, T. and Smith, B., 2001. Vague Reference and Approximating Judgments. *Spatial Cognition and Computation*, 3, 200–3.

Bittner, T. and Stell, J.G., 2000. Rough sets in approximate spatial reasoning. *In*: *Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing (RSCTC'2000) (LNCS)*. Springer-Verlag, 445–453.

Bittner, T. and Stell, J.G., 2003. Stratified Rough Sets and Vagueness. *In*: W. Kuhn, M. Worboys, and S. Timpf, eds. *COSIT 2003: Spatial Information Theory - Foundations of Geographic Information Science*. Berlin & Heidelberg: Springer-Verlag, 270–286.

Bucher, B., Clough, P., Finch, D., Joho, H., Purves, R., and Syed, A.K., 2005. Evaluation of SPIRIT Prototype Following Integration and Testing. Technical. University of Sheffield, UK.

Cai, G., 2007. Contextualization of Geospatial Database Semantics for Human—GIS Interaction. *Geoinformatica*, 11 (2), 217–237.

Cai, G., Wang, H., and Maceachren, A.M., 2003. Communicating Vague Spatial Concepts in Human-GIS Interactions: A Collaborative Dialogue Approach. *In*: *Proceedings of Conference on Spatial Information Theory (COSIT 2003)*. Kartause Ittingen: Berlin: Springer, 304–319.

Chen, J., Li, C., Li, Z., Li, C.L.Z., and Gold, C., 2003. Voronoi-based 9-Intersection Model for Topological Spatial Relations. *International Journal of Geographical Information Science*, 15 (3), 201–220.

Clementini, E., 2011. Directional relations and frames of reference. *GeoInformatica*, 1–21.

Clementini, E. and Di Felice, P., 1996. An algebraic model for spatial objects with indeterminate boundaries. *In*: P. Burrough and A. Frank, eds. *Geographic Objects with Indeterminate Boundaries*. London: Taylor & Francis, 155–169.

Clementini, E. and Di Felice, P., 1997. Approximate topological relations. *International Journal of Approximate Reasoning*, 16 (2), 173 – 204.

Clementini, E. and Di Felice, P., 2000. Spatial Operators. *ACM SIGMOD Rec*, 29 (3), 31–38.

Clementini, E., Di Felice, P., and Hernández, D., 1997. Qualitative Representation of Positional Information. *Artificial Intelligence*, 95, 317–356.

Clementini, E., Di Felice, P., and Van Oosterom, P., 1993. A Small Set of Formal Topological Relationships Suitable for End-User Interaction. *In*: *Third International Symposium on Advances On Spatial Databases (SSD'93)*. Singapore: Springer Berlin, Germany, 277–295.

Clough, P. and Pasley, R., 2010. Images and perceptions of neighbourhood extents. *In*: *Proceedings of the 6th Workshop on Geographic Information Retrieval*. New York, NY, USA: ACM, 13:1–13:2.

Clough, P., Sanderson, M., and Joho, H., 2004. Extraction of Semantic Annotations from Textual Web Pages. Technical. University of Sheffield, UK.

Cohn, A., 1995. A hierarchical representation of qualitative shape based on connection and convexity. *In*: A. Frank and W. Kuhn, eds. *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin / Heidelberg, 311–326.

Cohn, A. and Gotts, N., 1996. The "Egg-Yolk" representation of regions with indeterminate boundaries. *In*: P. Burrough and A. Frank, eds. *Geographic Objects with Indeterminate Boundaries*. Bristol, PA: Taylor and Francis, 171–187.

Cohn, A.G., Bennett, B., Gooday, J., and Gotts, N.M., 1997. Representing And Reasoning With Qualitative Spatial Relations About Regions. *In*: O. Stock, ed. *Spatial and Temporal Reasoning*. Kluwer, 97–134.

Cohn, A.G. and Hazarika, S.M., 2001. Qualitative Spatial Representation and Reasoning: An Overview. *Fundamenta Informaticae*, 46, 1–29.

Couclelis, H., Golledge, R.G., Gale, N., and Tobler, W., 1987. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology*, 7 (2), 99 – 122.

Davies, C., Holt, I., Green, J., Harding, J., and Diamond, L., 2009. User Needs and Implications for Modelling Vague Named Places. *Spatial Cognition & Computation*, 9 (3), 174 – 194.

Delboni, T.M., Laender, A.H.F., Borges, K.A.V., and Davis Jr., C.A., 2007. Semantic Expansion of Geographic Web Queries Based on Natural Language Positioning Expressions. *Transactions in GIS*, 11 (3), 377–397.

Dilo, A., De By, R.A., and Stein, A., 2007. A system of types and operators for handling vague spatial objects. *International Journal of Geographical Information Science*, 21 (4), 397–426.

Dini, L. and Di Tomaso, V., 1995. Events and individual in italian dynamic locative expressions. *In*: *Proceedings of the 4th Conference on the Cognitive Science of Natural Language Processing*.

Du, S., Wang, Q., and Yang, Y., 2004. Fuzzy description of fuzzy direction relations and their similarities. *In*: *Proceedings of 12th International Conference on Geoinformatics - Geospatial Information Research: Bridging the Pacific and Atlantic*. Gävle, Sweden, 496–502.

Dutta, S., 1990. Qualitative spatial reasoning: a semi-quantitative approach using fuzzy logic. *In*: *Proceedings of the first symposium on Design and implementation of large spatial databases*. New York, NY, USA: Springer-Verlag New York, Inc., 345–364.

Dutta, S., 1991. Approximate spatial reasoning: Integrating qualitative and quantitative constraints. *International Journal of Approximate Reasoning*, 5 (3), 307–330.

Egenhofer, M. and Franzosa, R., 1991. Point-set topological spatial relations. *International journal of geographical information systems*, 5 (2), 161–174.

Egenhofer, M. and Herring, J., 1990. Categorizing Binary Topological Relations between Regions, Lines and Point in Geographic Databases. Technical. Department of Surveying Engineering, University of Maine, Orono.

Egenhofer, M. and Mark, D., 1995. Naive Geography. *In*: Frank, A.U. and Kuhn, W., eds. *Spatial Information Theory: A Theoretical Basis for GIS*, LNCS 998, Berlin: Springer-Verlag, 1–15.

Eschenbach, C. and Kulik, L., 1997. An Axiomatic Approach to the Spatial Relations Underlying Left-Right and in Front of-Behind. *In*: G. Brewka, C. Habel, and B. Nebel, eds. *KI-97: Advances in Artificial Intelligence*. Berlin: Springer, 207–218.

Fisher, P., 1996. Boolean and fuzzy regions. *In*: P. Burrough and A. Frank, eds. *Geographic Objects with Indeterminate Boundaries*. London: Taylor & Francis, 87–94.

Fonte, C.C. and Lodwick, W., 2004. Areas of fuzzy geographical entities. *International Journal of Geographical Information Science*, 18 (2), 127–150.

Frank, A.U., 1992. Qualitative spatial reasoning about distances and directions in geographic space. *Journal of Visual Languages & Computing*, 3 (4), 343–371.

Freksa, C., 1992. Using Orientation Information for Qualitative Spatial Reasoning. *In*: A. Frank, I. Campari, and U. Formentini, eds. *Theories and Methods of SpatioTemporal Reasoning in Geographic Space, LNCS*. Springer-Verlag, 162–178.

Fu, G., Jones, C., and Abdelmoty, A., 2005. Ontology-Based Spatial Query Expansion in Information Retrieval. *In*: R. Meersman and Z. Tari, eds. *On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE*. Springer Berlin / Heidelberg, 1466–1482.

Gahegan, M., 1995. Proximity operators for qualitative spatial reasoning. *In*: A. Frank and W. Kuhn, eds. *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin / Heidelberg, 31–44.

Galton, A. and Hood, J., 2005. Anchoring: A New Approach to Handling Indeterminate Location in GIS. *In*: A.G. Cohn and D.M. Mark, eds. *Spatial Information Theory*. Springer Berlin Heidelberg, 1–13.

Gold, C.M., Nanter, J., and Yang, W., 1996. Outside-in: an alternative approach to forest map digitizing. *International journal of geographical information systems*, 10 (3), 291–310.

Goodchild, M. and Hill, L., 2006. Summary Report-Digital Gazetteer Research & Practice Workshop.

Goodchild, M. and Hill, L., 2008. Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22 (10), 1039–1044.

Grabowski, J. and Weiss, P., 1996. The prepositional inventory of languages: A factor that affects comprehension of spatial prepositions. *Language Sciences*, 18 (1–2), 19–35.

Grossman, D.A. and Frieder, O., 2004. *Information Retrieval: Algorithms and Heuristics (The Information Retrieval Series*. 2nd ed. Springer.

Grothe, C. and Schaab, J., 2009. Automated Footprint Generation from Geotags with Kernel Density Estimation and Support Vector Machines. *Spatial Cognition Computation*, 9 (3), 195–211.

Guarino, N., 1998. Formal Ontology and Information Systems. *In*: N. Guarino, ed. *Formal Ontology in Information Systems*. Trento, Italy: IOS Press, 3–15.

Guo, Q., Liu, Y., and Wieczorek, J., 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science*, 22 (10), 1067–1090.

Hastings, J.T., 2008. Automated Conflation of Digital Gazetteer Data. *International Journal of Geographical Information Science*, 22 (10), 1109–1127.

Hernández, D., 1991. Relative Representation of Spatial Knowledge. *In*: D. Mark and A. Frank, eds. *Cognitive and Linguistic Aspects of Geographic Space*. Dordrecht: Kluwer Academic, 373–385.

Hernández, D., Informatik, F.F., Clementini, E., and Felice, P.D., 1995. Qualitative Distances. *In*: A. Frank and W. Kuhn, eds. *Spatial Information Theory: A Theoretical Basis for GIS*. Berlin: Springer Verlag, 31–44.

Herskovits, A., 1985. Semantics and pragmatics of locative expressions. *Cognitive Science*, 9 (3), 341–378.

Hill, L., 2000. Core Elements of Digital Gazetteers: Placenames, Categories, and Footprints. *In*: J. Borbinha and T. Baker, eds. *Research and Advanced Technology for Digital Libraries*. Springer Berlin / Heidelberg, 280–290.

Hill, L., 2006. *Georeferencing: The Geographic Associations of Information*. Cambridge: MIT Press.

Hill, L., Frew, J., and Zheng, Q., 1999. Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library. *D-Lib Magazine*, 5 (1).

Hirtle, S. and Jonides, J., 1985. Evidence of hierarchies in cognitive maps. *Memory & Cognition*, 13 (3), 208–217.

Hollenstein, L. and Purves, R., 2010. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, 1 (1), 21–48.

Hood, J. and Galton, A., 2006. Implementing anchoring. *In*: *Proceedings of the 4th international conference on Geographic Information Science*. Berlin, Heidelberg: Springer-Verlag, 168–185.

Jaffe, A., Naaman, M., Tassa, T., and Davis, M., 2006. Generating summaries and visualization for large collections of geo-referenced photographs. *In*: *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*. New York, NY, USA: ACM, 89–98.

Janee, G., 2006. Rethinking Gazetteers and Interoperability. *In*: *Proceedings of the International Workshop on Digital Gazetteer Research & Practice*. Santa Barbara, California.

Janowicz, K., 2006. Sim-DL: Towards a semantic similarity measurement theory for the description logic ALCNR in geographic information retrieval. *In*: R. Meersman, Z. Tari, P. Herrero, and et al., eds. *SeBGIS 2006, OTM Workshops 2006. Lecture Notes in Computer Science*. Berlin: Springer, 1681–1692.

Janowicz, K. and Keßler, C., 2008. The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science*, 22 (10), 1129–1157.

Janowicz, K., Keßler, C., Schwarz, M., Wilkes, M., Panov, I., Espeter, M., and Bäumer, B., 2007. Algorithm, Implementation and Application of the SIM-DL Similarity Server. *In*: *Proceedings of Second International Conference on GeoSpatial Semantics, GeoS 2007*. Springer-Verlag Berlin Heidelberg, 128–145.

Janowicz, K., Raubal, M., and Kuhn, W., 2011. The Semantics of Similarity in Geographic Information Retrieval. *Journal of Spatial Information Science*, 1 (2), 29–57.

Jones, C., Alani, H., and Tudhope, D., 2001. Geographical Information Retrieval with Ontologies of Place. *In*: D. Montello, ed. *Spatial Information Theory*. Springer Berlin / Heidelberg, 322–335.
Jones, C.B., Purves, R., Ruas, A., Sanderson, M., Sester, M., Van Kreveld, M., and Weibel, R., 2002. Spatial information retrieval and geographical ontologies: An overview of the SPIRIT project. *In*: *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 387–388.
Jones, C.B. and Purves, R.S., 2008. Geographical information retrieval. *International Journal of Geographical Information Science*, 22 (3), 219–228.
Jones, C.B., Purves, R.S., Clough, P.D., and Joho, H., 2008. Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science*, 22 (10), 1045–1065.
Keßler, C., Janowicz, K., and Bishr, M., 2009. An agenda for the next generation gazetteer: geographic information contribution and retrieval. *In*: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York, NY, USA: ACM, 91–100.
Keßler, C., Maué, P., Heuer, J.T., and Bartoschek, T., 2009. Bottom-Up Gazetteers: Learning from the Implicit Semantics of Geotags. *In*: K. Janowicz, M. Raubal, and S. Levashkin, eds. *GeoSpatial Semantics*. Springer Berlin Heidelberg, 83–102.
Klippel, A. and Montello, D.R., 2007. Linguistic and nonlinguistic turn direction concepts. *In*: S. Winter, M. Duckham, L. Kulik, and B. Kuipers, eds. *Proceedings of the 8th international conference on Spatial information theory (COSIT'07)*. Springer-Verlag, 354–372.
Kuhn, W., 2001. Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15 (7), 613–631.
Kulik, L., 2001. A Geometric Theory of Vague Boundaries Based on Supervaluation. *In*: *Proceedings of the International Conference on Spatial Information Theory: Foundations of Geographic Information Science*. Presented at the COSIT 2001, London, UK, UK: Springer-Verlag, 44–59.
Landau, B. and Jackendoff, R., 1993. "What" and "where" in Spatial Language and Spatial Cognition. *Behavioral and Brain Sciences*, 16 (02), 217–238.
Lehmann, F. and Cohn, A.G., 1994. The EGG/YOLK reliability hierarchy: semantic data integration using sorts with prototypes. *In*: *Proceedings of the third international conference on Information and knowledge management*. Gaithersburg, MD: ACM, 272–279.
Leidner, J., 2007. Toponym Resolution in Text: Annotation, Evaluation and Applications of Spatial Grounding of Place Names. PhD Thesis. University of Edinburgh, Institute for Communicating and Collaborative Systems, School of Informatics.
Leidner, J.L., Sinclair, G., and Webber, B., 2003. Grounding spatial named entities for information extraction and question answering. *In*: *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 31–38.
Leveling, J. and Hartrumpf, S., 2008. On metonymy recognition for geographic information retrieval. *International Journal of Geographical Information Science*, 22 (3), 289–299.
Levinson, S.C., 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge University Press.

Li, H., Srihari, R.K., Niu, C., and Li, W., 2003. InfoXtract location normalization: a hybrid approach to geographic references in information extraction. *In*: *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 39–44.

Ligozat, G., 1993. Qualitative triangulation for spatial reasoning. *In*: A. Frank and I. Campari, eds. *Spatial Information Theory: A Theoretical Basis for GIS*. Springer Berlin / Heidelberg, 54–68.

Liu, Y., Guo, Q.H., Wieczorek, J., and Goodchild, M.F., 2009. Positioning localities based on spatial assertions. *International Journal of Geographical Information Science*, 23 (11), 1471–1501.

Longley, P.A., Goodchild, M., Maguire, D.J., and Rhind, D.W., 2010. *Geographic Information Systems and Science*. 3rd ed. John Wiley & Sons.

Lutz and Klien, E., 2006. Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20 (3), 233–260.

Mark, D. and Egenhofer, M., 1994. Calibrating the Meanings of Spatial Predicates from Natural Language: Line-Region Relations. Presented at the Sixth International Symposium on Data Handling, Edinburgh, Scotland: London: Taylor & Francis, 538–553.

Martins, B., 2011. Delimiting imprecise regions with georeferenced photos and land coverage data. *In*: *Proceedings of the 10th international conference on Web and wireless geographical information systems*. Berlin, Heidelberg: Springer-Verlag, 219–229.

Meegan, R. and Mitchell, A., 2001. It's Not Community Round Here, It's Neighbourhood': Neighbourhood Change and Cohesion in Urban Regeneration Policies. *Urban Studies*, 38 (12), 2167–2194.

Montello, D.R., Goodchild, M.F., Gottsegen, J., and Fohl, P., 2003. Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries. *Spatial Cognition & Computation*, 3 (2-3), 185–204.

Munoz, A. and Moguerza, J.M., 2006. Estimation of high-density regions using one-class neighbor machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (3), 476 –480.

Overell, S. and Rüger, S., 2008. Using co-occurrence models for placename disambiguation. *Int. J. Geogr. Inf. Sci.*, 22 (3), 265–287.

Purves, R., Clough, P., and Joho, H., 2005. Identifying imprecise regions for geographic information retrieval using the web. *In*: *Proceedings of GIS Research UK 13th Annual Conference*. 313–318.

Ragni, M., Fangmeier, T., and Schleipen, S., 2007. What about negation in spatial reasoning? *In*: *Proceedings of the 28th Annual Cognitive Science Conference (CogSci 07)*. Lawrence Erlbaum Associates Inc., 1409 – 1414.

Randell, D.A., Cui, Z., and Cohn, A.G., 1992. A Spatial Logic based on Regions and Connection. *In*: *Proceedings of 3rd International Conference on Principles of Knowledge, Representation and Reasoning (KR'92)*. Morgan Kaufmann, 165–176.

Rattenbury, T.Y.E. and Naaman, M.O.R., 2009. Methods for Extracting Place Semantics from Flickr Tags. *ACM Transactions on the Web*, 3 (1), 1–30.

Rauch, E., Bukatin, M., and Baker, K., 2003. A confidence-based framework for disambiguating geographic terms. *In*: *Proceedings of the HLT-NAACL 2003 workshop on Analysis of geographic references - Volume 1*. Stroudsburg, PA, USA: Association for Computational Linguistics, 50–54.

Retz-Schmidt, G., 1988. Various Views on Spatial Prepositions. *AI Magazine*, 9 (2), 95.

Robinson, V.B., 1990. Interactive machine acquisition of a fuzzy spatial relation. *Computers & Geosciences*, 16 (6), 857–872.

Sanderson, M. and Kohler, J., 2004. Analyzing Geographic Queries. *In*: *Proceedings of Workshop on Geographic Information Retrieval SIGIR*.

Schleipen, S., Ragni, M., and Fangmeier, T., 2007. Negation in Spatial Reasoning. *In*: J. Hertzberg, M. Beetz, and R. Englert, eds. *KI 2007: Advances in Artificial Intelligence*. Springer Berlin / Heidelberg, 175–189.

Schlieder, C., 1995. Reasoning about ordering. *In*: A. Frank and W. Kuhn, eds. *Spatial Information Theory A Theoretical Basis for GIS*. Springer Berlin / Heidelberg, 341–349.

Schlieder, C., Vögele, T., and Visser, U., 2001. Qualitative Spatial Representation for Information Retrieval by Gazetteers. *In*: D. Montello, ed. *Spatial Information Theory*. Springer Berlin / Heidelberg, 336–351.

Schockaert, S., 2011. Vague regions in Geographic Information Retrieval. *SIGSPATIAL Special*, 3 (2), 24–28.

Schockaert, S., Cock, M.D., and Kerre, E.E., 2005. Automatic acquisition of fuzzy footprints. *In*: *Proceedings of the Workshop on Semantic-based Geographical Information Systems, LNCS 3762*. Presented at the ACM SIGIR, Agia Napa: Springer, 1077–1086.

Schockaert, S., Smart, P.D., Abdelmoty, A.I., and Jones, C.B., 2008. Mining Topological Relations from the Web. *In*: *Proceedings of the 19th International Conference on Database and Expert Systems Application*. Washington, DC, USA: IEEE Computer Society, 652–656.

Schockaert, S., Smart, P.D., and Twaroch, F.A., 2011. Generating approximate region boundaries from heterogeneous spatial information: An evolutionary approach. *Information Sciences*, 181 (2), 257–283.

Shariff, A.R., Egenhofer, M.J., and Mark, D.M., 1998. Natural-language spatial relations between linear and areal objects: the topology and metric of English-language terms. *International Journal of Geographical Information Science*, 12 (3), 215–246.

Sharma, J., Flewelling, D.M., and Egenhofer, M.J., 1994. A Qualitative Spatial Reasoner. *In*: T.C. Waugh and R.G. Healey, eds. *Proceedings of the 6th International Symposium on Spatial Data Handling (SDH'94)*. 665–681.

Smith, B., 1995. On Drawing Lines on a Map. *In*: A. Frank and W. Kuhn, eds. *Spatial Information Theory*. Springer Berlin, 475–484.

Smith, B. and Varzi, A.C., 1997. Fiat and Bona Fide Boundaries. *In*: *Philosophy and Phenomenological Research*. 401–420.

Stokes, N., Li, Y., Moffat, A., and Rong, J., 2008. An empirical study of the effects of NLP components on Geographic IR performance. *International Journal of Geographical Information Science*, 22 (3), 247–264.

Talmy, L., 1983. How language structures space. *In*: H. Pick and L. Acredolo, eds. *Spatial Orientation: Theory, Research, and Application*. New York: Plenum Press, 225–282.

Tenbrink, T. and Freksa, C., 2009. Contrast Sets in Spatial and Temporal Language. *Cognitive Processing*, 10 (Supplement 2), 322–324.

Tobler, W., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (2), 234–240.

Tversky, B., 1993. Cognitive Maps, Cognitive Collages and Spatial Mental Models. *In*: A. Frank and I. Campari, eds. *Spatial Information Theory. A Theoretical Basis for GIS*. Springer Berlin, 14–24.

Twaroch, F.A., Jones, C.B., and Abdelmoty, A.I., 2008. Acquisition of a vernacular gazetteer from web sources. *In*: *Proceedings of the first international workshop on Location and the web*. New York, NY, USA: ACM, 61–64.

Twaroch, F.A., Smart, P.D., and Jones, C.B., 2008. Mining the web to detect place names. *In*: *Proceedings of the 2nd international workshop on Geographic information retrieval*. New York, NY, USA: ACM, 43–44.

Vögele, T., Schlieder, C., and Visser, U., 2003. Intuitive Modelling of Place Name Regions for Spatial Information Retrieval. *In*: W. Kuhn, M. Worboys, and S. Timpf, eds. *Spatial Information Theory. Foundations of Geographic Information Science*. Springer Berlin / Heidelberg, 239–252.

Vorwerg, C. and Rickheit, G., 1998. Typicality Effects in the Categorization of Spatial Relations. *In*: *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. London, UK, UK: Springer-Verlag, 203–222.

Wang, F., 1994. Towards a natural language user interface: an approach of fuzzy query. *International journal of geographical information systems*, 8 (2), 143–162.

Wang, L., Wang, C., Xie, X., Forman, J., Lu, Y., Ma, W.-Y., and Li, Y., 2005. Detecting dominant locations from search queries. *In*: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. New York, NY, USA: ACM, 424–431.

Wieczorek, J., Guo, Q., and Hijmans, R., 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, 18 (8), 745–767.

Williamson, T., 1994. *Vagueness*. London: Routledge.

Winter, S. and Truelove, M., 2010. Talking about Place Where It Matters. *In*: A. Frank and D. Mark, eds. *Las Navas 20th Anniversary Meeting on Cognitive and Linguistic Aspects of Geographic Space*. Las Navas, Spain.

Worboys, M., Duckham, M., and Kulik, L., 2004. Commonsense Notions of Proximity and Direction in Environmental Space. *Spatial Cognition & Computation*, 4 (4), 285–312.

Worboys, M.F., 2001. Nearness relations in environmental space. *International Journal of Geographical Information Science*, 15 (7), 633–651.

Yao, X. and Thill, J.-C., 2006. Spatial queries with qualitative locations in spatial information systems. *Computers, Environment and Urban Systems*, 30 (4), 485 – 502.